





GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 11, 2002, 15:52:53 ; Search time 43 Seconds  
(without alignments)  
2.093 Million cell updates/sec

Title: us-09-880-457-4

Perfect score: 455  
Sequence: 1 MAILFLSLQLILLIPSISH.....TSAGTLQLQLWNCNLRHH 87

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.1  
Ygapop 10.0 , Ygapext 0.1  
Fgapop 6.0 , Fgapext 0.1  
Delop 6.0 , Delext 0.1

Searched: 6 seqs, 517263 residues

Total number of hits satisfying chosen parameters: 12

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=soft -Q=us-09-880-457-4 -DB=deberry457.seq  
-SUFFIX=ptc -OUT=align4\_hits -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits  
-START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200  
-THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFWT=ptc  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -NCPU=6 -NO\_XLPXY  
-NEG\_SCORES=0 -LONGLOG -THREADS=1 -XGAPOP=10 -XGAPEXT=0.1 -FGAPOP=6  
-FGAPEXT=0.1 -YGAPOP=10 -YGAPEXT=0.1 -DELOP=6 -DELEXT=0.1

Database : deberry457.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	440	96.7	1186	1 abk33576	TOIG of: abk3357
2	440	96.7	1186	1 abk35882	TOIG of: abk35882
3	440	96.7	1186	1 ax362321	TOIG of: ax362321
4	434	95.4	165707	1 ac104794	TOIG of: ac10479
5	434	95.4	190748	1 ac010969	TOIG of: ac01096
6	117.9	25.9	157250	1 ac027264	TOIG of: ac02726
7	117.9	25.9	190748	1 ac010969	TOIG of: ac01096
8	114.8	25.2	157250	1 ac027264	TOIG of: ac02726
9	113.2	24.9	165707	1 ac104794	TOIG of: ac10479
10	101.1	22.2	1186	1 abk33576	TOIG of: abk3357
11	101.1	22.2	1186	1 abk35882	TOIG of: abk35882
12	101.1	22.2	1186	1 ax362321	TOIG of: ax36232

#### ALIGNMENTS

RESULT 1  
abk33576  
; TOIG of: abk33576 check: 121 from: 1 to: 1186  
; ID ABK33576 standard; CDNA; 1186 BP.

ABK33576;

08-MAY-2002 (first entry)

CDNA encoding human PRO protein, Seq ID No 81.

Human; secreted protein; PRO; tumour; lung cancer; colon cancer;  
breast cancer; prostate tumour; rectal tumour; liver tumour;  
pericyte proliferation; chondrocyte cell proliferation;  
tumour necrosis factor-alpha; gene; ss.

Homo sapiens.

WO200208288-A2.

31-JAN-2002.

29-JUN-2001; 2001WO-US21066.

20-JUL-2000; 2000US-219556P.

25-JUL-2000; 2000US-220585P.

25-JUL-2000; 2000US-220605P.

25-JUL-2000; 2000US-220607P.

25-JUL-2000; 2000US-220624P.

25-JUL-2000; 2000US-220638P.

25-JUL-2000; 2000US-220664P.

25-JUL-2000; 2000US-220666P.

26-JUL-2000; 2000US-220893P.

28-JUL-2000; 2000WO-US20710.

23-AUG-2000; 2000WO-US23522.

24-AUG-2000; 2000WO-US23328.

15-SEP-2000; 2000US-000000P.

10-NOV-2000; 2000WO-US30873.

28-NOV-2000; 2000US-253646P.

01-DEC-2000; 2000WO-US32678.

20-DEC-2000; 2000US-0747259.

20-DEC-2000; 2000WO-US34956.

28-FEB-2001; 2001WO-US06520.

10-MAY-2001; 2001US-0854280.

25-MAY-2001; 2001WO-US17092.

(GETH ) GENENTECH INC.

Baker KP, Desnoyers L, Gerritsen ME, Goddard A, Godowski PJ;

Grimaldi JC, Gurney AL, Smith V, Stephan JF, Watanabe CK, Wood WI;

WPI; 2002-172001/22.

P-PSDB; AAU83632.

One hundred and twenty two nucleic acids encoding PRO polypeptides,  
useful for treating a PRO related disorder and for diagnosing tumours  
such as lung cancer, colon cancer, breast tumour, prostate tumour, rectal  
tumour or liver tumour -

Claim 2; Figure 81; 359pp; English.

The invention relates to one hundred and twenty two nucleic acids  
encoding PRO polypeptides. The sequences of the 122 PRO polynucleotides  
encode human secreted proteins. The PRO nucleic acids, polypeptides,  
agonists and antagonists are useful for treating a PRO related disorder.  
The PRO polypeptides are useful for diagnosing tumours, especially lung  
cancer, colon cancer, breast tumour, prostate tumour, rectal tumour or  
liver tumour. The PRO polypeptides are useful for stimulating the  
proliferation of, or gene expression, in pericyte cells, for stimulating  
the proliferation or differentiation of chondrocyte cells, for  
stimulating the release of tumour necrosis factor-alpha from human blood,  
for stimulating or inhibiting the proliferation of normal human dermal  
fibroblast cells. The PRO polypeptide may also be used as molecular  
weight markers and for tissue typing. The PRO nucleic acids have  
applications in molecular biology, including use as hybridisation probes,  
and in chromosome and gene mapping. ABK33536-ABK33657 represent human  
PRO protein coding sequences of the invention.

```

DB:
us-09-880-457-4 (1-87) x ax358828 (1-1186)

Qy 1 MetAlaIleLeuThrLeuSerLeuGlnLeuLeuLeuLeuLeuProSerIleSerHis 20
Dbb 447 ATGGCAATCCTGAGCCTCAACTCATCTTGTTATTATAATCAATATCCCAT 506
Qy 21 GluAlaHisLysThrSerLeuSerSerTrpLysHisAspGlnAspTrpAlaAsnValSer 40
Dbb 507 GAGGCTCATAAACAGAGCTCTTCTCTTGGAACATGACCAAGATGGGCAACGCTCC 566
Qy 41 AsnMetThrPheSerAsnGlyLysLeuArgValLysGlyIleTyrTrpArgAsnAlaAsp 60
Dbb 567 AACATGACTTTCAGCAACGGAAACTAAGAGTCAAGGCATTATTACCGAATGCCGAC 626
Qy 61 IleCysSerArgHisArgValThrSerAlaGlyLeuThrLeuGlnAspLeuGlnLeuTrp 80
Dbb 627 ATTTCGCTCGACATCCGCTAACCTCAGCAGGCGTAACCTCTGAGGACCTTCAGCTATGG 686
Qy 81 CysAsnLeuArgIleIle 86
Dbb 687 TGTAATTTGAGGTCAGTG 704

RESULT 3
ax362321
; TOIG of: ax362321 check: 121 from: 1 to: 1186
;
; LOCUS AX362321 1186 bp DNA linear PAT 15-F
; DEFINITION Sequence 81 from Patent WO0208288.
; ACCESSION AX362321
; VERSION AX362321.1 GI:18694618
; KEYWORDS
; SOURCE
; ORGANISM Homo sapiens
; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
;
; REFERENCE
; 1 Baker,K.P., Desnoyers,L., Gerritsen,M.E., Goddard,A.,
; Godowski,P.J., Grimaldi,J.C., Gurney,A.L., Smith,V., Stephan
; Watanabe,C.K. and Wood,W.I.
; TITL Secretd and transmembrane polypeptides and nucleic acids en
; the same
; JOURNAL Patent: WO 0208288-A 81 31-JAN-2002;
; Genentech, Inc. (US)
; FEATURES
; source
; location/Qualifiers
; /organism="Homo sapiens"
; /db_xref="taxon:9606"
; BASE COUNT 314 a 285 c 299 g 288 t
; ORIGIN
;
;
; AX362321 Length: 1186 December 11, 2002 15:47 Type: N Check: 121
ax362321

Alignment Scores:
Pred. No.: 0.211 Length: 1186
Score: 440.00 Matches: 84
Percent Similarity: 98.84% Conservative: 1
Best Local Similarity: 97.67% Mismatches: 1
Query Match: 96.70% Indels: 0
DB: 1 Gaps: 0

us-09-880-457-4 (1-87) x ax362321 (1-1186)

Qy 1 MetAlaIleLeuThrLeuSerLeuGlnLeuLeuLeuLeuProSerIleSerHis 20
Dbb 447 ATGGCAATCCTGAGCCTCAACTCATCTTGTTATTATAATCAATATCCCAT 506
Qy 21 GluAlaHisLysThrSerLeuSerSerTrpLysHisAspGlnAspTrpAlaAsnValSer 40
Dbb 507 GAGGCTCATAAACAGAGCTCTTCTCTTGGAACATGACCAAGATGGGCAACGCTCC 566

```

```

QY 41 AsnMetThrPheSerAsnGlyLysLeuArgValLysGlyLeuTyrArgAsnAlaasp 60
|||||
Db 567 AACATGACITTCAGCAACGGAACAACTAAGAGTCAAGGCAATTTATTACCGAATCCGCAC 626
|||||
QY 61 IleCysSerArgHisArgValThrSerAlaGlyLeuThrLeuGlnAspLeuGlnLeuTrp 80
|||||
Db 627 ATTGCTCTCGACATCGGTACCTCAGCAGGCGCTAACTCTCGCAGGACCTTCAGCTATGG 686
|||||
QY 81 CysAsnLeuArgIleIle 86
|||||
Db 687 TGTAAATTTGAGGTCAAGTG 704
|||||
RESULT 4
ac104794
; TOIG of: ac104794 check: 7810 from: 1 to: 165707
;
; LOCUS AC104794 165707 bp DNA linear HTG 30-APR-2002
; DEFINITION Homo sapiens chromosome 2 clone RP11-254F7, WORKING DRAFT SEQUENCE,
; 3 unordered pieces.
; ACCESSION AC104794
; VERSION AC104794.3 GI:20340520
; KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
; SOURCE Homo sapiens.
; ORGANISM Homo sapiens
; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
; REFERENCE 1 (bases 1 to 165707)
; AUTHORS Waterston,R.H.
; TITLE The sequence of Homo sapiens clone
; JOURNAL Unpublished
; REFERENCE 2 (bases 1 to 165707)
; AUTHORS Waterston,R.H.
; TITLE Direct Submission
; JOURNAL Submitted (21-DEC-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
; REFERENCE 3 (bases 1 to 165707)
; AUTHORS Waterston,R.H.
; TITLE Direct Submission
; JOURNAL Submitted (30-APR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
; COMMENT On Apr 30, 2002 this sequence version replaced gi:19339129.
;
; ----- Genome Center -----
; Center: Washington University Genome Sequencing Center
; Center code: WUGSC
; Web site: http://genome.wustl.edu/gsc/index.shtml
; Contact: submissions@wustl.edu
; ----- Project Information -----
; Center project name: H.NH0254F07
; ----- Summary Statistics -----
; Sequencing vector: M13; 0%
; Chemistry: Dye-primer ET; 100%
; Chemistry: Dye-terminator Big Dye; 100% of reads
; Assembly program: Phrap; version 0.990319
; Consensus quality: 162662 bases at least Q40
; Consensus quality: 163189 bases at least Q30
; Consensus quality: 164786 bases at least Q20
; Insert size: 9479; agarose-fp
; Quality coverage: 12.78 in Q20 bases; agarose-fp
; Quality coverage: 10.53 in Q20 bases; sum-of-contigs
;
; * NOTE: This is a 'working draft' sequence. It currently
; * consists of 3 contigs. The true order of the pieces
; * is not known and their order in this sequence record is
; * arbitrary. Gaps between the contigs are represented as
; * runs of N, but the exact sizes of the gaps are unknown.
; * This record will be updated with the finished sequence
; * as soon as it is available and the accession number will

```

```

* be preserved.
;
; 1112: contig of 1112 bp in length
; 1113: gap of unknown length
; 1213: contig of 75015 bp in length
; 76227: gap of unknown length
; 76228: contig of 89380 bp in length.
; 76328: Location/Qualifiers
; FEATURES
; 1..165707
; /organism="Homo sapiens"
; /db_xref="taxon:9606"
; /chromosome="2"
; /clone="RP11-254F7"
; 1..1112
; /note="assembly_name:Contig21"
; 1213..76227
; /note="assembly_name:Contig68"
; 76328..165707
; /note="assembly_name:Contig69"
; BASE COUNT 42887 a 42242 c 39083 g 41295 t 200 others
; ORIGIN
;
; AC104794 Length: 165707 December 11, 2002 15:47 Type: N Check: 7810
; ac104794
;
; Alignment Scores:
; Pred. No.: 0.229 Length: 165707
; Score: 434.00 Matches: 83
; Percent Similarity: 97.67% Conservative: 1
; Best Local Similarity: 96.51% Mismatches: 2
; Query Match: 95.38% Indels: 0
; DB: 1 Gaps: 0
;
; us-09-880-457-4 (1-87) x ac104794 (1-165707)
;
; QY 1 MetAlaIleLeuThrLeuSerLeuGlnLeuLeuLeuLeuLeuProSerIleSerHis 20
|||||
Db 41120 ATGGCAATCCTGATGCTCAGCCTTCAACTCATCTTGTATTATACCATATCCAT 41179
|||||
QY 21 GluAlaHisLysThrSerLeuSerSerTrpLysHisAspGlnAspTrpAlaAsnValSer 40
|||||
Db 41180 GAGGCTCATAAACGAGTCTTCTTCTTGGAAACATGACCAAGATTGGCAACGCTCC 41239
|||||
QY 41 AsnMetThrPheSerAsnGlyLysLeuArgValLysGlyLeuTyrArgAsnAlaasp 60
|||||
Db 41240 AACATGACTTTCAGCAACGGAACAACTAAGAGTCAAGGCAATTTATTACCGAATCCGCAC 41299
|||||
QY 61 IleCysSerArgHisArgValThrSerAlaGlyLeuThrLeuGlnAspLeuGlnLeuTrp 80
|||||
Db 41300 ATTGCTCTCGACATCGGTACCTCAGCAGGCGCTAACTCTCGCAGGACCTTCAGCTATGG 41359
|||||
QY 81 CysAsnLeuArgIleIle 86
|||||
Db 41360 TGTAAATTTGAGGTCAAGTG 41377
|||||
RESULT 5
ac10969/c
; TOIG of: ac10969 check: 475 from: 1 to: 190748
;
; LOCUS AC010969 190748 bp DNA linear PRI 07-NOV-2001
; DEFINITION Homo sapiens BAC clone RP11-95D17 from 2, complete sequence.
; ACCESSION AC010969
; VERSION AC010969.11 GI:13677120
; KEYWORDS HTG.
; SOURCE Homo sapiens.
; ORGANISM Homo sapiens
; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
; REFERENCE 1 (bases 1 to 190748)
; AUTHORS Sulston,J.E. and Waterston,R.
; TITLE Toward a complete human genome sequence
; JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
; MEDLINE 99063792
; PUBMED 9847074

```



```

; repeat_region 12467..12594
; /rpt_family="Alu"
; repeat_region 12596..12678
; /rpt_family="L1"
; repeat_region 12698..12954
; /rpt_family="Alu"
; repeat_region 12955..13065
; /rpt_family="Alu"
; repeat_region 13194..13433
; /rpt_family="L2"
; repeat_region 13557..13667
; /rpt_family="Alu"
; misc_feature 13614..14110
; /note="similar to EST BF770132 (NID:gl2118032)"
; misc_feature 13623..14122
; /note="similar to EST BF770135 (NID:gl2118035)"
; misc_feature 13630..14128
; /note="similar to EST BF770134 (NID:gl2118034)"
; misc_feature 13635..14128
; /note="similar to EST BF770127 (NID:gl2118027)"
; misc_feature 13688..13942
; /note="similar to EST BF770793 (NID:gl2118693)"
; misc_feature 13727..14127
; /note="similar to EST BF770796 (NID:gl2118696)"
; misc_feature 13729..14126
; /note="similar to EST BF770791 (NID:gl2118691)"
; repeat_region 13735..13801
; /rpt_family="MIR"
; misc_feature 14062..14259
; /note="similar to EST AL545813 (NID:gl2878338)"
; misc_feature 14144..14259
; /note="similar to EST BF699834 (NID:gl11985242)"
; misc_feature 14205..14259
; /note="similar to EST BF668191 (NID:gl11942086)"
; misc_feature 14229..14259
; /note="similar to EST AL571368 (NID:gl2928594)"
; misc_feature 14422..14474

Alignment Scores:
Pred. No.: 0.229 Length: 190748
Score: 434.00 Matches: 83
Percent Similarity: 97.67% Conservative: 1
Best Local Similarity: 96.51% Mismatches: 2
Query Match: 95.38% Indels: 0
DB: 1 Gaps: 0

us-09-880-457-4 (1-87) x ac010969 (1-190748)

Qy 1 MetAlaIleLeuThrLeuSerLeuGlnLeuIleLeuLeuLeuIleProSerIleSerHis 20
Db 174485 ATGCCAATCTGATGCTCAGCCTTCACATCATCTGTTATTATACCATCAATATCCCAT 174426

Qy 21 GluAlaHisLysThrSerLeuSerSerTrpLysHisAspGlnAspTrpAlaAsnValSer 40
Db 174425 GAGGCTCATAAAACGAGTCTTTCTTCTGGAAACATGACCAAGATTGGGCAACGCTCTCC 174366

Qy 41 AsnMetThrPheSerAsnGlyLysLeuArgValLysGlyIleTyrTrpArgAsnAlaasp 60
Db 174365 AACATGACTTTCAGCAACGAAACATTAAGAGCAAGGCAATTTATTACCGAATCCGAC 174306

Qy 61 IleCysSerArgHisArgValThrSerAlaGlyLeuThrLeuGlnAspLeuGlnLeuTrp 80
Db 174305 ATTGTCTCTGACATCGCGTAACCTCAGCAGCGCTTAACCTCTCAGGACCTTCAGCTATGG 174246

Qy 81 CysAsnLeuArgIleIle 86
Db 174245 TGTAATTTGAGTCAAGTG 174228

RESULT 6
ac027264/c
; TOIG of: ac027264 check: 9979 from: 1 to: 157250
; LOCUS AC027264 157250 bp DNA linear HTG 24-AUG-2002

```

```

; DEFINITION Homo sapiens chromosome 2 clone RP11-164P8 map 2. WORKING DRAFT
; ACCESSION AC027264.2 GI:10567976
; VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
; KEYWORDS human.
; SOURCE
; ORGANISM
; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
; REFERENCE 1 (bases 1 to 157250)
; AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
; TITLE Homo sapiens chromosome 2, clone RP11-164P8
; JOURNAL Unpublished
; REFERENCE 2 (bases 1 to 157250)
; AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
; Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
; Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
; Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
; Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
; Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
; Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
; Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
; Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
; Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
; Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
; McCarthy,M., McEwan,P., McGurk,A., McKernan,K., MCPheeters,K.,
; Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
; Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
; O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
; Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
; Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
; Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
; Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
; Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
; Young,G., Zainoun,J., Zimmer,A. and Zody,M.
; TITLE Direct Submission
; JOURNAL Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome
; Research, 320 Charles Street, Cambridge, MA 02141, USA
; REFERENCE 3 (bases 1 to 157250)
; AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
; Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavskiy,L.,
; Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
; Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
; DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
; FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
; Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
; Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
; Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,
; Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
; MCPheeters,K., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
; Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
; O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K.,
; Pierre,N., Pisanli,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
; Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
; Sounges,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
; Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
; Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo.A.,
; Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
; Zimmer,A. and Zody,M.
; TITLE Direct Submission
; JOURNAL Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
; Research, 320 Charles Street, Cambridge, MA 02141, USA
; COMMENT On Oct 4, 2000 this sequence version replaced gi:7331634.
; All repeats were identified using RepeatMasker:
; Smit, A.F.A. & Green, P. (1996-1997)
; http://ftp.genome.washington.edu/RM/RepeatMasker.html
; ----- Genome Center
; Center: Whitehead Institute/ MIT Center for Genome Research
; Center code: WIBR
; Web site: http://www-seq.wi.mit.edu
; Contact: sequence_submissions@genome.wi.mit.edu
; ----- Project Information
; Center project name: L8993

```





align4\_hits

Wed Dec 11 15:55:47 2002

```

Db 89177 CATATCAGCATCCAAGAGAGAAATTTCAAGCATGCAGTTAGATGATAAAGTCTCCAGTTCA 89118
QY 53 -----GlyIleTyrArg----- 57
                                     :|||||
Db 89117 GGAGAGGCTGACCGCGGACACATAATTTGGAAGCTGTTAGTTTATATATAATCAATCATG 89058
QY 58 -----AsnAlaAspIleCysSer---ArgHis----- 65
                                     :|||||
Db 89057 AAATCTAACATGCTATTGTTGTAACATACGCCATTATTTAAAGTACTGCTAACAAGAA 88998
QY 66 -ArgVal--ThrSerAlaGlyLeuThrLeuGln----- 75
                                     :|||||
Db 88997 AAGATGCTACAAATTAATTTGACATGCCAGTCATTTGAAGATAAATCTTGATTCTTA 88938
QY 76 -AspLeuGlnLeuTrpCys-----AsnLeuArgIleIle 86
                                     :|||||
Db 88937 GGATGTTAAATGTTGTTATGGGAGGGACATTTTAAATTTCAAGATAGTA 88887

RESULT 7
ac010969
; TOIG of: ac010969 check: 475 from: 1 to: 190748
;
; LOCUS AC010969 190748 bp DNA linear PRI 07-NOV-2001
; DEFINITION Homo sapiens BAC clone RP11-95D17 from 2, complete sequence.
; ACCESSION AC010969
; VERSION AC010969.11 GI:13677120
; KEYWORDS HTG.
; SOURCE
; ORGANISM Homo sapiens.
; Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
; REFERENCE 1 (bases 1 to 190748)
; AUTHORS Sulston, J.E. and Waterston, R.
; TITLE Toward a complete human genome sequence
; JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
; MEDLINE 99063792
; PUBMED 9847074
; REFERENCE 2 (bases 1 to 190748)
; AUTHORS Sun, H., Abbott, A. and Le, T.P.
; TITLE The sequence of Homo sapiens BAC clone RP11-95D17
; JOURNAL Unpublished
; REFERENCE 3 (bases 1 to 190748)
; AUTHORS Waterston, R.H.
; TITLE Direct Submission
; JOURNAL Submitted (28-SEP-1999) Genome Sequencing Center, Washington
; REFERENCE 4 (bases 1 to 190748)
; AUTHORS University School of Medicine, 4444 Forest Park Parkway, St. Louis,
; MO 63108, USA
; REFERENCE 5 (bases 1 to 190748)
; AUTHORS Waterston, R.H.
; TITLE Direct Submission
; JOURNAL Submitted (20-APR-2001) Genome Sequencing Center, Washington
; REFERENCE 6 (bases 1 to 190748)
; AUTHORS University School of Medicine, 4444 Forest Park Parkway, St. Louis,
; MO 63108, USA
; REFERENCE 7 (bases 1 to 190748)
; AUTHORS Waterston, R.
; TITLE Direct Submission
; JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington
; University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
; COMMENT On Apr 19, 2001 this sequence version replaced gi:11128441.
; ----- Genome Center
; Center: Washington University Genome Sequencing Center
; Center code: WUGSC
; Web site: http://genome.wustl.edu/gsc
; Contact: sapiens@watson.wustl.edu
; ----- Summary Statistics
; -----
; Center project name: H_NH0095D17
; -----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION: This clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION: The RP11-95D17 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenoe, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION: The clone sequenced to the left is RP11-521D12. Actual start of this clone is at base position 1 of RP11-95D17; actual end is at base position 190748 of RP11-95D17.

There are polymorphic base differences between RP11-95D17 and the redundant clone AC062035. Data from AC062035 was used to finish RP11-95D17.

Unresolved tandem repeat from base position 181200 to 187300. Size information from restriction digest suggests that the full repeat may not be represented.

FEATURES	Source	Location/Qualifiers
	1. .190748	/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="2"
		/map="2"
		/clone="RP11-95D17"
		/clone.lib="RP11-11"
repeat_region	185..230	/rpt_family="L2"
repeat_region	325..805	/rpt_family="L1"
repeat_region	806..1114	/rpt_family="Alu"
repeat_region	1449..1567	/rpt_family="MIR"
repeat_region	2175..2369	/rpt_family="MIR"
repeat_region	2626..3646	/rpt_family="L2"
repeat_region	3716..3745	/rpt_family="AT-rich"
repeat_region	3798..4101	/rpt_family="Alu"
repeat_region	4196..5255	/rpt_family="L1"
repeat_region	5256..5559	/rpt_family="Alu"
repeat_region	5560..6453	/rpt_family="L1"
repeat_region	6939..6989	

```

repeat_region /rpt_family="MIR"
7701.8007
repeat_region /rpt_family="Alu"
8242.8532
repeat_region /rpt_family="MER1_type"
8533.8834
repeat_region /rpt_family="Alu"
8835.9300
repeat_region /rpt_family="MER1_type"
9322.9647
repeat_region /rpt_family="L2"
9652.9784
repeat_region /rpt_family="MaLR"
9784.9848
repeat_region /rpt_family="Alu"
9862.10061
repeat_region /rpt_family="L1"
10062.10360
repeat_region /rpt_family="Alu"
10361.10646
repeat_region /rpt_family="L1"
10647.10948
repeat_region /rpt_family="Alu"
10949.11090
repeat_region /rpt_family="L1"
11135.11420
repeat_region /rpt_family="Alu"
11424.11704
repeat_region /rpt_family="L1"
11705.12015
repeat_region /rpt_family="Alu"
12016.12045
repeat_region /rpt_family="L1"
12047.12327
repeat_region /rpt_family="Alu"
12335.12416
repeat_region /rpt_family="L1"
12467.12594
repeat_region /rpt_family="Alu"
12596.12678
repeat_region /rpt_family="L1"
12698.12954
repeat_region /rpt_family="Alu"
12955.13065
repeat_region /rpt_family="Alu"
13194.13433
repeat_region /rpt_family="L2"
13557.13667
repeat_region /rpt_family="Alu"
13614.14110
misc_feature /note="similar to EST BF770132 (NID:g12118032)"
13623.14122
misc_feature /note="similar to EST BF770135 (NID:g12118035)"
13630.14128
misc_feature /note="similar to EST BF770134 (NID:g12118034)"
13635.14128
misc_feature /note="similar to EST BF770127 (NID:g12118027)"
13688.13942
misc_feature /note="similar to EST BF770793 (NID:g12118693)"
13727.14127
misc_feature /note="similar to EST BF770796 (NID:g12118696)"
13729.14126
repeat_region /note="similar to EST BF770791 (NID:g12118691)"
13735.13801
misc_feature /rpt_family="MIR"
14062.14259
misc_feature /note="similar to EST AL545813 (NID:g12878338)"
14144.14259
misc_feature /note="similar to EST BF699834 (NID:g11985242)"
14205.14259
misc_feature /note="similar to EST BF668191 (NID:g11942086)"
14229.14259
misc_feature /note="similar to EST AL571368 (NID:g12928594)"

```

```

; misc_feature 14422.14474
Alignment Scores:
Score: 9.14 Length: 150748
Pred. No.: 117.90 Matches: 38
Percent Similarity: 18.33% Conservative: 19
Best Local Similarity: 12.22% Mismatches: 23
Query Match: 25.91% Indels: 231
DB: 1 Gaps: 0
us-09-880-457-4 (1-87) x ac010969 (1-190748)
Qy 7 SerLeuGln-----Leulle-----LeuLeuLeuProSerile 18
Db 30747 AGTGTGCAGGGGAATGCCACTTTTAAACAGTCAGATCTCATGAGAACTCCCTCACTA 30806
Qy 19 SerHisGluAlaHis-----LysThr-----S 26
Db 30807 TCACAAGAACAGCATGGGGGAACTGCCCCATGATCAGTCACCTCCACCTAGTCCCT 30866
Qy 26 erLeuSerSerTrpLys-----HisAspGlnAspTrpAlaAsnValSerAsnMetTh 43
Db 30867 CCCTCCAGACATGGGATTACAAATTCAGATTTGGGTGGGACACAGACCAAC 30926
Qy 43 rPheSerAsnGlyLys-----LeuArgValLys-----52
Db 30927 CATATCAGCATCCAAAGAGGAAATTTCAAGCATGCAGTTAGATGTAAAGTCTGCAGTTC 30986
Qy 53 -----GlyIleTyrTyrArg-----57
Db 30987 GGAGAGAGTCTGAGCCGGAGACATAAATTTCCGAAGCTGTAGTTTATATATAATCATG 31046
Qy 58 -----AsnAlaAspIleCysSer---ArgHis-----65
Db 31047 AAATCTAACATGCTATTGTTGTAACATACGCCCATATTTTAAGTACTGCTAAGAAAGAA 31106
Qy 66 -ArgVal--ThrSerAlaGlyLeuThrLeuGln-----75
Db 31107 AAGAGTCTACAATTCATTTTGGACATGCCAGTGTGTAAGATAAATCTTGATTCTTA 31166
Qy 76 -AspLeuGlnLeuTrpCys-----AsnLeuArgIleIle 86
Db 31167 GGATGTTAAATGTGTATGGTGGAGGGGACATTTTAAATTTGAAGATAGTA 31217
RESULT 8
ac027264
; TOIG of: ac027264 check: 9979 from: 1 to: 157250
; LOCUS AC027264 157250 bp DNA linear HTG 24-AUG-2002
; DEFINITION Homo sapiens chromosome 2 clone RP11-164P8 map 2, WORKING DRAFT
; SEQUENCE, 21 unordered pieces.
; ACCESSION AC027264
; VERSION AC027264.2 GI:10567976
; KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
; SOURCE human.
; ORGANISM Homo sapiens
; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
; REFERENCE 1 (bases 1 to 157250)
; AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
; TITLE Homo sapiens chromosome 2, clone RP11-164P8
; JOURNAL Unpublished
; REFERENCE 2 (bases 1 to 157250)
; AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
; Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
; Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
; Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
; Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
; Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
; Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
; Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
; Howland,J.C., Iliev,I., Johnson,I., Jones,C., Kann,L., Karatas,A.,
; Klein,J., LaRocque,K., Lamazares,R., Landers,R., Lehoczy,J.,

```



```

; /note="assembly_fragment"
; 56844..64404
; /note="assembly_fragment"
; 64505..71871
; /note="assembly_fragment"
; 71972..82019
; /note="assembly_fragment"
; 82120..107411
; /note="assembly_fragment"
; 107512..116168
; /note="assembly_fragment"
; 116269..128458
; /note="assembly_fragment"
; 128559..146712
; /note="assembly_fragment"
; 146813..157250
; /note="assembly_fragment"
; clone_end:T7
; vector_side:right
; BASE COUNT 45395 a 32097 c 31766 g 45991 t 2001 others
; ORIGIN
; AC027264 Length: 157250 December 11, 2002 15:49 Type: N Check: 9979
ac027264

Alignment Scores:
Pred. No.: 9.31 Length: 157250
Score: 114.80 Matches: 43
Percent Similarity: 14.03% Conservative: 20
Best Local Similarity: 9.58% Mismatches: 24
Query Match: 25.23% Indels: 362
DB: 1 Gaps: 2

us-09-880-457-4 (1-87) x ac027264 (1-157250)

QY 1 MetAlaLeu-----ThrLeuSerLeuInLeuLeuLeuLeuLeu 15
Db 65340 ATGGTGTCTTAGACATATTTCGGCATATTAGCTTTACATGATATTATTCAGTTG 65399
|||||
QY 16 Pro-----SerIleSerHisGlu-AlaHis 23
Db 65400 CCTCTTTTATGAGAACATTAGAAACATGCTAATAAAACCCACCAGGAGGCACAC 65459
|||||
QY 24 -----LysThrSer-----LeuSerSerTr 30
Db 65460 AGCCTCTTAGACAGAGTCTGCTGTGAAGAGGATGGTCCCTTATTTAGTGGCATG 65519
|||||
QY 30 pLysHisAsp---Gln-----AspTrpAlaAsnVal----- 39
Db 65520 GAGCCATGAACCTTCAGGCTCTGCGACTGGAGCTCTGCTGCTTTGGATCTGATGTTGGC 65579
|||||
QY 39 ----- 39
Db 65580 TCCCTATTGTTGCAAGAGGTGTCATTATGTTTGGCTGAAGATGCTATCAGCAGAG 65639
|||||
QY 39 ----- 39
Db 65640 TGCTTCTCAGACCTCTCCTTTTGTCTTCTTGATGATGAGGATTTGGAATCAGACAACT 65699
|||||
QY 40 -----Sera 41
Db 65700 CAGCCTAAGGCGAGTCAGGGCTCTGAGCGAGGGGCTAAGGGGCTCTCTGAGAAAAA 65759
|||||
QY 41 snMetThrPheSerAsngly----- 47
Db 65760 GGCTGAGTTTTAGTCAGGCTCTCTGTGGCATTTTACTGCCAGCTGAGCCCTTTTGTGC 65819
|||||
QY 47 ----- 47
Db 65820 TTCCTCTGCAGTTCTAGGCTCTGGTCTAGAACAGGTAAGGTTACACAGATAAGTTTGTGC 65879
|||||
QY 47 ----- 47

```

```

Db 65880 CTAAAGTCATGTCATCACTCACTGAAACTTCATACAGTTAGAAATTTAGTTTATCTACTTCATGCA 65939
QY 48 -----LysLeuArgVal-LysGlyIle 54
Db 65940 GAGAAAAAATCTCAGCAGTTTCATCGGTGAGTATACATAGCTCTATGATATACTACTG 65999
QY 55 TyrTyr-----ArgAsnAlaAspIleCys-----Sera 64
Db 66000 CATTATACCTGAGTATTAAATCTGAGATAATCCACCAATCCACCTCTCCCTGTGTAATA 66059
QY 64 rGHs-----ArgValThrSer-----AlaGlyLeuThrLeuGln----- 75
Db 66060 AGCAAAAGGTACGATCAGCAGATTAAGCTGCTCAGGAGCAAACTCTGCAGCTGGAGGAGG 66119
QY 76 -----AspLeuGlnLeuTrpCys----- 81
Db 66120 CAGGGCATGGCACACCATGCTTCTCCTCTCAGGATAAAATCTGAAATCATTGACAT 66276
QY 82 -----AsnLeuArgIleIle--His 87
Db 66180 TATAACTCTTTCTCTCTCTCAGGATAAAATCTGAAATCATTGACAT 66276

RESULT 9
ac104794/c
; TOIG Of: ac104794 check: 7810 from: 1 to: 165707
; LOCUS AC104794 165707 bp DNA linear HTG 30-APR-2002
; DEFINITION Homo sapiens chromosome 2 clone RP11-254F7, WORKING DRAFT SEQUENCE,
; 3 unordered pieces.
; ACCESSION AC104794
; VERSION AC104794.3 GI:20340520
; KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
; SOURCE Homo sapiens.
; ORGANISM Homo sapiens
; Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
; Waterston, R.H.
; TITLE The sequence of Homo sapiens clone
; JOURNAL Unpublished
; REFERENCE 2 (bases 1 to 165707)
; AUTHORS Waterston, R.H.
; TITLE Direct Submission
; JOURNAL Submitted (21-DEC-2001) Genome Sequencing Center, Washington
; University School of Medicine, 4444 Forest Park Parkway, St. Louis,
; MO 63108, USA
; REFERENCE 3 (bases 1 to 165707)
; AUTHORS Waterston, R.H.
; TITLE Direct Submission
; JOURNAL Submitted (30-APR-2002) Genome Sequencing Center, Washington
; University School of Medicine, 4444 Forest Park Parkway, St. Louis,
; MO 63108, USA
; COMMENT
; On Apr 30, 2002 this sequence version replaced gi:19339129.
; ----- Genome Center
; Center: Washington University Genome Sequencing Center
; Center code: WUGSC
; Web site: http://genome.wustl.edu/gsc/index.shtml
; Contact: submissions@watson.wustl.edu
; ----- Project Information
; Center project name: HNH0254F07
; ----- Summary Statistics
; Sequencing vector: pM3; 0%
; Sequencing vector: plasmid; 100%
; Chemistry: Dye-primer ET; 0% of reads
; Chemistry: Dye-terminator Big Dye; 100% of reads
; Assembly program: Phrap; version 0.990319
; Consensus quality: 162662 bases at least Q40
; Consensus quality: 163189 bases at least Q30
; Consensus quality: 164786 bases at least Q20
; Insert size: 9479; agarose-fp
; Insert size: 167795; sum-of-contigs

```

align4\_hits

Wed Dec 11 15:55:47 2002

Quality coverage: 12.78 in Q20 bases; agarose-fp  
Quality coverage: 10.53 in Q20 bases; sum-of-contrigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1112: contig of 1112 bp in length  
1113 1212: gap of unknown length  
1213 76227: contig of 75015 bp in length  
76228 76327: gap of unknown length  
76328 165707: contig of 89380 bp in length.  
-----  
FEATURES  
Location/Qualifiers  
1..165707  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/clone="RP11-254F7"  
1..1112  
/note="assembly\_name:Contig21"  
1213..76227  
/note="assembly\_name:Contig68"  
76328..165707  
/note="assembly\_name:Contig69"  
42887 a 42242 c 39083 g 41295 t 200 others  
BASE COUNT  
ORIGIN

AC104794 Length: 165707 December 11, 2002 15:47 Type: N Check: 7810  
ac104794

Alignment Scores:  
Pred. No.: 9.4 Length: 165707  
Score: 113.20 Matches: 46  
Percent Similarity: 18.75% Conservative: 14  
Best Local Similarity: 14.38% Mismatches: 26  
Query Match: 24.88% Indels: 237  
DB: 1 Gaps: 1  
us-09-880-457-4 (1-87) x ac104794 (1-165707)  
Qy 2 Ala1Leu-----ThrLeuSerLeuGln-----Leu-----I 11  
|||:|||||  
Db 53674 GCCTTTTGAATGGTCTGCTGCTCAGCAGACATTTGCCCTGCAACACCTAGTCAGGA 53615  
Qy 11 leLeuLeuLeuLeu-----ProSerIleSerHisGluAla----- 22  
|||||:|||||:|  
Db 53614 TTCTCTACTGGTATCAAGTACAAACTCCAGTGACACTTCAGAGAGACGCTGACTAAC 53555  
Qy 22 ----- 22  
Db 53554 TAGGAGACCTCATTTACTAGTACTTACGGGGCCAGCCAGATCATTTGGGGAAGATGGAGTCT 53495  
Qy 23 -----HisLysThrSerLeuSerSer-- 29  
Db 53494 CTAACCTACTAAATTAATAAATTTGGATGGCATGCTTACACTTCACCTCATTTGTCAGATT 53435  
Qy 30 -----TrpLysHisAsp-----GlnA 35  
Db 53434 CTTGATATCTTCTGAAGTCAATGGGTCATGAGTGTGGAATGTAATATTTAATGAGG 53375  
Qy 35 spTrpAlaAsnValSer-----AsnMetThrPhe-----S 45  
Db 53374 ACA--GCTGAGGTGAAATTTCTTCCAAATATGTTTACACATTTACATTTAGCCCTTT 53317  
Qy 45 exAsnGlyLysLeuArg-----ValLysGlyIleTyrTyrArgAsnAlaAsp11 61  
Db 53316 CCAATGA-AAACTAAATAGCTCTGCTGTCGCAAAATCTTCTGTGGATATGCGCTCTGT 53258  
Qy 61 eCys-----SerArgHisArgValThrSerAlaG1 71

Db 53257 GTGTGTTTGAAGATACACTTTTCCCTTTTGATTGATTAAGACATTCAGTTGATTTTACTGG 53198  
Qy 71 YLeuThrLeuGlnAspLeuGlnLeu-----Trp----- 80  
Db 53197 AAATACACA-CAACAAATTCCTTTTACATAAACTGGGATTTCTTAGGACAGGCAGCGTGCCT 53139  
Qy 81 --CysAsnLeu-ArgIleIleHis 87  
Db 53138 GTTGTATCTTACAGAGTGTACAC 53115  
RESULT 10  
abk33576/c  
; TOIG of: abk33576 check: 121 from: 1 to: 1186  
; ID ABK33576 standard; cDNA; 1186 BP.  
; AC ABK33576;  
; DT 08-MAY-2002 (first entry)  
; DE CDNA encoding human PRO protein, Seq ID No 81.  
; KW Human; secreted protein; PRO; tumour; lung cancer; colon cancer;  
; KW breast cancer; prostate tumour; rectal tumour; liver tumour;  
; KW pericyte cell proliferation; chondrocyte cell proliferation;  
; KW tumour necrosis factor-alpha; gene; ss.  
; OS Homo sapiens.  
; PN WO200208288-A2.  
; PD 31-JAN-2002.  
; PF 29-JUN-2001; 2001WO-US21066.  
; PR 20-JUL-2000; 2000US-219556P.  
; PR 25-JUL-2000; 2000US-220585P.  
; PR 25-JUL-2000; 2000US-220603P.  
; PR 25-JUL-2000; 2000US-220607P.  
; PR 25-JUL-2000; 2000US-220624P.  
; PR 25-JUL-2000; 2000US-220638P.  
; PR 25-JUL-2000; 2000US-220664P.  
; PR 25-JUL-2000; 2000US-220666P.  
; PR 26-JUL-2000; 2000US-220893P.  
; PR 28-JUL-2000; 2000WO-US20710.  
; PR 23-AUG-2000; 2000WO-US23522.  
; PR 24-AUG-2000; 2000WO-US23328.  
; PR 15-SEP-2000; 2000US-000000P.  
; PR 10-NOV-2000; 2000WO-US30873.  
; PR 28-NOV-2000; 2000US-253646P.  
; PR 01-DEC-2000; 2000WO-US32678.  
; PR 20-DEC-2000; 2000US-0747259.  
; PR 28-DEC-2000; 2000WO-US34956.  
; PR 10-FEB-2001; 2001WO-US06520.  
; PR 10-MAY-2001; 2001US-0854280.  
; PR 25-MAY-2001; 2001WO-US17092.  
; XX (GETH ) GENENTECH INC.  
; PA Baker KP, Desnoyers L, Gerritsen ME, Goddard A, Godowski PJ;  
; PI Grimaldi JC, Gurney AL, Smith V, Stephan JF, Watanabe CK, Wood WI;  
; XX WPI: 2002-172001/22.  
; DR P-PSDB; AAU83632.

One hundred and twenty two nucleic acids encoding PRO polypeptides,  
useful for treating a PRO related disorder and for diagnosing tumours  
such as lung cancer, colon cancer, breast tumour, prostate tumour, rectal  
tumour or liver tumour -  
Claim 2; Figure 81; 359pp; English.  
; XX

The invention relates to one hundred and twenty two nucleic acids encoding PRO polypeptides. The sequences of the 122 PRO polynucleotides encode human secreted proteins. The PRO nucleic acids, polypeptides, agonists and antagonists are useful for treating a PRO related disorder. The PRO polypeptides are useful for diagnosing tumours, especially lung cancer, colon cancer, breast tumour, prostate tumour, rectal tumour or liver tumour. The PRO polypeptides are useful for stimulating the proliferation of, or gene expression, in pericyte cells, for stimulating the release of tumour necrosis factor-alpha from human blood, for stimulating or inhibiting the proliferation of normal human dermal fibroblast cells. The PRO polypeptide may also be used as molecular weight markers and for tissue typing. The PRO nucleic acids have applications in molecular biology, including use as hybridisation probes, and in chromosome and gene mapping. ABK33536-ABK33657 represent human PRO protein coding sequences of the invention.

align4\_hits

Wed Dec 11 15:55:47 2002

|||||  
Db 245 TTGCTCAGAGTGCAGGAACACGGGGGAGAGAGCGGGGTAGCATCTTCCCGACAGCGTT 186  
QY 74 -----LeuGlnAspLeuGln----- 78  
Db 185 AGGAGGTACGTGAATGTCACACAGTAGAACAATCCCTGCAAAAATGACAGAGGTGCC 126  
QY 79 -----LeuTrp-----CysAsn----- 82  
Db 125 TCAGTAACATTTCTAGGGGACCAGCAATCTGGGGCATGTAAACAGATCCCTCCAGGGTGA 66  
QY 83 -----LeuArg-----IleIleHis 87  
Db 65 AGGACAAGTCTGCGTGGGCCCTCCACCCCTACTCTCCAACCTCCCTTCAC 15

Search completed: December 11, 2002, 15:54:29  
Job time : 96 secs

Db 125 TCAGTAACATTTCTAGGGGACCAGCAATCTGGGGCATGTAAACAGATCCCTCCAGGGTGA 66  
QY 83 -----LeuArg-----IleIleHis 87  
Db 65 AGGACAAGTCTGCGTGGGCCCTCCACCCCTACTCTCCAACCTCCCTTCAC 15  
RESULT 12  
ax362321/c  
; TOIG of: ax362321 check: 121 from: 1 to: 1186  
; LOCUS AX362321 1186 bp DNA linear PAT 15-FEB-2002  
; DEFINITION Sequence 81 from Patent WO0208288.  
; ACCESSION AX362321  
; VERSION AX362321.1 GI:18694618  
; KEYWORDS  
; SOURCE human.  
; ORGANISM Homo sapiens  
; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
; REFERENCE 1  
; AUTHORS Baker, K.P., Desnoyers, L., Gerritsen, M.E., Goddard, A.,  
; Godowski, P.J., Grimaldi, J.C., Gurney, A.L., Smith, V., Stephan, J.P.,  
; Watanabe, C.K., and Wood, W.I.  
; TITLE Secreted and transmembrane polypeptides and nucleic acids encoding  
; the same  
; JOURNAL Patent: WO 0208288-A 81 31-JAN-2002;  
; Genentech, Inc. (US)  
; FEATURES  
; source  
; 1..1186  
; /organism="Homo sapiens"  
; /db\_xref="taxon:9606"  
; BASE COUNT 314 a 285 c 299 g 288 t  
; ORIGIN  
; AX362321 Length: 1186 December 11, 2002 15:47 Type: N Check: 121  
; ax362321

Alignment Scores:  
Pred. NO.: 10 Length: 1186  
Score: 101.10 Matches: 43  
Percent Similarity: 16.09% Conservative: 18  
Best Local Similarity: 11.35% Mismatches: 23  
Query Match: 22.22% Indels: 297  
DB: 1 Gaps: 1  
us-09-880-457-4 (1-87) x ax362321 (1-1186)  
QY 4 LeuThrLeuSerLeuGlnLeuIle-----LeuLeuLeu-----IleProSerIleSerHi 20  
Db 601 TTGACTCTTAGTTTCCGTTGCTGAAGTCATGTTGGAGACGCTTGGCCCAATCTTGGTCA 542  
QY 20 sGluAla-----HisLysThrSerLeu-----SerSerTrpLysHisAsp----- 33  
Db 541 TGTTCCTCAAGAAAGAAAGACTGCTTTTATGAGCCTCTATGGGATATTGATGATTATAAAC 482  
QY 34 -----GlnAspTrpAlaAsnValSer-AsnMetThrPheSer 45  
Db 481 AAGATGAGTTGAGGCTGAGCGTCAGGATT--GCCATTCTTAGCAATACTTCTTCTCC 424  
QY 46 -----AsnGlyLys-----LeuArgValLysGly----- 53  
Db 423 CTATTAAATTCATCAGAACGATTCCTGAGCTGGCAGGCGCAACGATACACCTTCA 364  
QY 54 -----Ile-----TyrTyrArgAsnAla 59  
Db 363 GTGTCCCCCTCGGGCTCTCTGATTCCTCAGCCAGCACTCCTATGTGAGGGGIC 304  
QY 60 AspIleCysSerArg-----HisArgValThr----- 68  
Db 303 G--GTCGACAGCAACTGTGTGTCAGCTTCCAGCCACAGGCGCATCAGTCAGCCCATCCA 246  
QY 69 -----SerAlaGlyLeuThr----- 73

